PROCEEDINGS *of the* FOURTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY

Held at the Statistical Laboratory
University of California
June 20–July 30, 1960,
with the support of
University of California
National Science Foundation
Office of Naval Research
Office of Ordnance Research
Air Force Office of Research
National Institutes of Health

VOLUME IV

CONTRIBUTIONS TO BIOLOGY AND PROBLEMS OF MEDICINE

EDITED BY JERZY NEYMAN

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1961

ON GENERAL LAWS AND THE MEANING OF MEASUREMENT IN PSYCHOLOGY

G. RASCH
THE DANISH INSTITUTE OF EDUCATIONAL RESEARCH, COPENHAGEN

1. Introduction

Lately the statistical methods of psychometrics have been severely criticized in psychological quarters. Thus Skinner [1] maintains that if order is to to be found in human and animal behavior, then it should be extracted from investigations into individuals, and that psychometric methods are inadequate for such purposes since they deal with groups of individuals. And as regards abnormal psychology Zubin [2] states: "Recourse must be had to individual statistics, treating each patient as a separate universe. Unfortunately, present day statistical methods are entirely group-centered, so that there is a real need for developing individual-centered statistics."

In a recently published book [3] I have developed three models for reactions to certain attainment tests and intelligence tests. Within the very limited areas covered, these models represent an attempt to meet this challenge. In fact, each model specifies a distribution function for the potential responses of a given person to a given stimulus of a certain set of allied stimuli, and this distribution function depends upon a parameter characterizing the person and a parameter characterizing the stimulus. The models have a remarkable property in common that renders it possible, in the analysis of the data, to detach the personal parameters from the stimulus parameters, and vice versa. And furthermore, we may check the adequacy of the model itself independently of both sets of parameters.

The present paper is concerned with a rather large class of models sharing this *separability property*, and also with some of the implications of this type of models.

2. A model for tests in oral reading

Let me begin with a description of two of the above mentioned models which pertain to static situations, leaving the third one, which is dynamical, for another occasion.

A large number $\nu = 1, \cdots, n$, of children were given a few tests, $i = 1, \cdots, k$, in reading aloud and on each occasion the number of misreadings was counted.

For the sake of simplicity, we shall consider only the cases where all texts were completed.

For reasons not to be discussed here (compare chapter 2 in [3]) the number of misreadings $x_{\nu i}$ made by child number ν in text number i is assumed to follow a Poisson law,

$$P\{x_{\nu i}\} = e^{-\xi_{\nu i}} \frac{\xi_{\nu i}^{x_{\nu i}}}{x_{\nu i}!}. \tag{2.1}$$

We shall attempt to describe the parameter $\xi_{\nu i}$ as the product of two factors, one pertaining to the person and one to the text,

$$\xi_{\nu i} = \theta_\nu \sigma_i. \tag{2.2}$$

At this stage we might, in keeping with the usual psychometric reasoning, assume that the θ, the "underlying continuum of inabilities," follow some distribution. A normal distribution is usually preferred, but since its composition with the Poisson law would be hard to handle, a gamma distribution might be tolerated for once.

It is, however, a main point of the present paper that for models like (2.1) any assumption about populations in an ordinary psychometrical sense is superfluous as far as comparisons of tests and comparisons of persons go. In fact, if we assume stochastic independence of the k test results for the same child and also of the test results for different children, the distribution function for the whole set of observations $((x_{\nu i}))$, with $\nu = 1, \cdots, n$ and $i = 1, \cdots, k$, becomes

$$P\{((x_{\nu i}))\} = e^{-\theta.\sigma.} \frac{\prod_\nu \theta_\nu^{x_{\nu\cdot}} \prod_i \sigma_i^{x_{\cdot i}}}{\prod_\nu \prod_i x_{\nu i}!} \tag{2.3}$$

where a dot indicates summation over the corresponding index.

From (2.3) it is easy algebra to deduce the distribution of the grand total,

$$P\{x_{\cdot\cdot}\} = e^{-\theta.\sigma.} \frac{(\theta.\sigma.)^{x_{\cdot\cdot}}}{x_{\cdot\cdot}!} \tag{2.4}$$

and the two conditional distributions of the marginal sums, given the grand total,

$$P\{(x_{\nu\cdot})|x_{\cdot\cdot}\} = \binom{x_{\cdot\cdot}}{x_{1\cdot}, \cdots, x_{n\cdot}} \prod_\nu \left(\frac{\theta_\nu}{\theta.}\right)^{x_{\nu\cdot}}, \tag{2.5}$$

$$P\{(x_{\cdot i})|x_{\cdot\cdot}\} = \binom{x_{\cdot\cdot}}{x_{\cdot 1}, \cdots, x_{\cdot k}} \prod_i \left(\frac{\sigma_i}{\sigma.}\right)^{x_{\cdot i}}. \tag{2.6}$$

It is also easily realized that for given $x_{\cdot\cdot}$ the two sets of marginal sums are stochastically independent, that is,

$$P\{(x_{\nu\cdot}), (x_{\cdot i})|x_{\cdot\cdot}\} = P\{(x_{\nu\cdot})|x_{\cdot\cdot}\}P\{(x_{\cdot i})|x_{\cdot\cdot}\}. \tag{2.7}$$

Finally, (2.3) with (2.5) and (2.6) leads to

$$(2.8) \qquad P\{((x_{\nu i}))|(x_{\nu \cdot}), (x_{\cdot i})\} = \frac{\prod_{\nu} x_{\nu \cdot}! \prod_{i} x_{\cdot i}!}{x_{\cdot \cdot}! \prod_{\nu} \prod_{i} x_{\nu i}!}.$$

Formulas (2.5) and (2.6) give a quite clear cut separation of the parameters: *the ratios $\theta_\nu/\theta.$ are to be estimated from one marginal, the ratios $\sigma_i/\sigma.$ from the other one, and these estimators are sufficient.* With one of the tests chosen as reference, with $\sigma = 1$, all the other σ can be estimated uniquely. In consequence, $\theta.$ may be estimated from (2.4), and accordingly, the estimates of the θ_ν are fixed.

Formula (2.8) tells us that, given the two marginals, the total set of observations obeys a distribution which is independent of all parameters. *In consequence, we may check the model without involving any of the unknown parameters.* In order to indicate in which way (2.8) may be employed for such control purposes, we consider a special case of (2.6), namely,

$$(2.9) \qquad P\{x_{\nu 1}, \cdots, x_{\nu k}|x_{\nu \cdot}\} = \binom{x_{\nu \cdot}}{x_{\nu 1}, \cdots, x_{\nu k}} \prod_{i} \left(\frac{\sigma_i}{\sigma.}\right)^{x_{\nu i}}$$

from which it follows that the expected value of any $x_{\nu i}$ for given total $x_{\nu \cdot}$ is proportional to $x_{\nu \cdot}$,

$$(2.10) \qquad E[x_{\nu i}|x_{\nu \cdot}] = \frac{\sigma_i}{\sigma.} x_{\nu \cdot}.$$

Thus, if we collect children into groups according to their total number of misreadings in the k tests all together and, for each group, average the number of misreadings in test number i, these averages should, apart from random variations which may be judged by a χ^2-test, increase proportionally to $x_{\nu \cdot}$.

This test, which is, by the way, passed beautifully by my observations, may be conceived as an application of (2.8) in which we, without specifying explicitly any alternative to our model (2.1) and (2.2), are looking for trouble in a more or less definite direction, namely, for the possibility that the relative difficulties of the tests may vary with $x_{\nu \cdot}$, that is, with the reading inability of the children.

3. A model for a type of intelligence test

For the responses to the items of an intelligence test the following model is suggested. (For the background of this choice, see chapter 5 in [3].) The probability that person number ν gives a correct answer to item number i is

$$(3.1) \qquad \frac{\theta_\nu \sigma_i}{1 + \theta_\nu \sigma_i},$$

assumed independent of the answers to the preceding questions.

For the following algebra it is convenient to reformulate this model in terms of another random variable $a_{\nu i}$, defined as 1 if the answer of person ν to item i is correct, 0 if it is not correct. In both cases we have

$$P\{a_{\nu i}\} = \frac{(\theta_\nu \sigma_i)^{a_{\nu i}}}{1 + \theta_\nu \sigma_i}. \tag{3.2}$$

Under the assumption of independence between the answers to different items for any given person, as well as between the answers of different persons, the probability of the whole set of answers

$$A = (a_{\nu i}), \qquad \nu = 1, \cdots, n; i = 1, \cdots, k, \tag{3.3}$$

of the n persons to the k items becomes

$$P\{A\} = \frac{\prod_\nu \theta_\nu^{a_{\nu\cdot}} \prod_i \sigma_i^{a_{\cdot i}}}{\gamma(\Theta, \Sigma)}, \tag{3.4}$$

where the denominator is

$$\gamma(\Theta, \Sigma) = \prod_\nu \prod_i (1 + \theta_\nu \sigma_i). \tag{3.5}$$

Since $P\{A\}$ is the same for all matrices (3.3) with the same two marginal vectors

$$\mathfrak{A}_{*\cdot} = (a_{1\cdot}, \cdots, a_{n\cdot}), \qquad \mathfrak{A}_{\cdot *} = (a_{\cdot 1}, \cdots, a_{\cdot k}), \tag{3.6}$$

the joint distribution of these vectors is

$$P\{\mathfrak{A}_{*\cdot}, \mathfrak{A}_{\cdot *}\} = \begin{bmatrix} \mathfrak{A}_{*\cdot} \\ \mathfrak{A}_{\cdot *} \end{bmatrix} \frac{\prod_\nu \theta_\nu^{a_{\nu\cdot}} \prod_i \sigma_i^{a_{\cdot i}}}{\gamma(\Theta, \Sigma)} \tag{3.7}$$

where the coefficient on the right denotes the number of all such possible matrices with all $a_{\nu i} = 1$ or 0. On summing (3.7) over all $\mathfrak{A}_{*\cdot}$, we obtain the distribution

$$P\{\mathfrak{A}_{*\cdot}\} = \frac{\gamma_{\mathfrak{A}_{*\cdot}}(\Sigma)}{\gamma(\Theta, \Sigma)} \prod_\nu \theta_\nu^{a_{\nu\cdot}}, \tag{3.8}$$

where

$$\gamma_{\mathfrak{A}_{*\cdot}}(\Sigma) = \sum_{\mathfrak{A}_{\cdot *}} \begin{bmatrix} \mathfrak{A}_{*\cdot} \\ \mathfrak{A}_{\cdot *} \end{bmatrix} \sigma_1^{a_{\cdot 1}} \cdots \sigma_k^{a_{\cdot k}}. \tag{3.9}$$

Now, by dividing (3.8) into (3.7) we obtain the conditional distribution of one marginal vector, given the other,

$$P\{\mathfrak{A}_{\cdot *} | \mathfrak{A}_{*\cdot}\} = \begin{bmatrix} \mathfrak{A}_{*\cdot} \\ \mathfrak{A}_{\cdot *} \end{bmatrix} \frac{\prod_i \sigma^{a_{\cdot i}}}{\gamma_{\mathfrak{A}_{*\cdot}}(\Sigma)}, \tag{3.10}$$

and a symmetrical argument yields the symmetrical formula

$$P\{\mathfrak{A}_{*\cdot} | \mathfrak{A}_{\cdot *}\} = \begin{bmatrix} \mathfrak{A}_{*\cdot} \\ \mathfrak{A}_{\cdot *} \end{bmatrix} \frac{\prod_\nu \theta_\nu^{a_{\nu\cdot}}}{\gamma_{\mathfrak{A}_{*\cdot}}(\Theta)}. \tag{3.11}$$

Finally, we may divide (3.7) into (3.4), thus obtaining the conditional dis-

tribution of the total matrix of observations, given the two marginal vectors,

$$P\{A|\mathfrak{A}_{*.}, \mathfrak{A}_{.*}\} = \frac{1}{\begin{bmatrix}\mathfrak{A}_{*.}\\ \mathfrak{A}_{.*}\end{bmatrix}}. \tag{3.12}$$

Formulas (3.10), (3.11), and (3.12) form the analogues to (2.5), (2.6), and (2.8), being, however, somewhat weaker since $\mathfrak{A}_{*.}$ and $\mathfrak{A}_{.*}$ for given $a_{..}$ do not appear to be stochastically independent. The main feature, however, has been retained.

On the basis of (3.10) *we may estimate the item parameters independently of the personal parameters, the latter having been replaced by something observable, namely, by the individual total number of correct answers. Furthermore, on the basis of* (3.11) *we may estimate the personal parameters without knowing the item parameters which have been replaced by the total number of correct answers per item. Finally,* (3.12) *allows for checks on the model* (3.2) *which are independent of all of the parameters, relying only upon the observations.*

4. Models with separate parameters

The practical results of applying these two models have been encouraging enough to justify a theoretical study of their basic properties as a preliminary to attempts at covering much wider fields of psychology. Formally, at least, we shall confine ourselves to the types of experiments or observations where all persons in question have been exposed to the same set of stimuli belonging to a certain, potentially large, class of allied stimuli. And at present we shall deal only with single individuals in static situations, assuming that preceding responses of a given person have no influence upon later responses, and the responses of any given person are unaffected by the responses of any other person. Situations to which these conditions do not apply may be considered under a dynamic point of view, such as processes of learning or of adaptation, or as group psychological phenomena.

In passing it may be noted that this assumption of independence of items is conditional to a given "ability." It is, therefore, not at variance with the well-established high intercorrelations between items which are produced under our model by the strong variation of the "abilities" in the "population" considered.

The responses may, of course, be quantitative, such as reaction times, or estimated sizes or velocities. However, they may just as well be qualitative, such as answers to a questionnaire or solutions to problems in an intelligence test, which are recorded as, for example, correct, incorrect, or not answered.

In psychometric practice the specification of the statistical model for the latter type of data is usually preceded by a so-called quantification or scoring of the qualitative observations, which is quite often nothing more than an enumeration of possible responses. In view of Dirichlet's definition of a function as just a correspondence between two sets of elements, such procedures would seem unnecessary as prerequisites to mathematical formulations of general laws. In

the present approach we shall, therefore, start from the original observations themselves, quantitative or qualitative as the case may be.

In accordance with the two models already considered we shall attempt to characterize each individual ν by a parameter θ_ν which, however, is not necessarily one-dimensional. And similarly, each stimulus i will be characterized by a parameter σ_i which also may be of a higher dimension. To each combination of individual and stimulus we shall assign a probability distribution on the set of possible responses. For simplicity we shall on this occasion consider only the case of a finite number m of possible responses, to be denoted $x^{(1)}, \cdots, x^{(m)}$. The probability that individual ν gives the response $x_{\nu i} = x$, which, by the way, also may be a vector, to the stimulus i is assumed to be determined by x and a parameter $\xi_{\nu i}$, which is, in turn, determined entirely by θ_ν and σ_i,

$$\xi_{\nu i} = \mu(\theta_\nu, \sigma_i). \tag{4.1}$$

This formulation allows for the possibility that the characteristics of persons and stimuli show some sort of random variation. In this case θ_ν and σ_i are interpreted as parameters in probability distributions.

In section 5 we shall refer to this representation; at present it suffices to keep in mind that the probability that $x_{\nu i} = x$ is conditioned by θ_ν and σ_i.

In the proofs of the separability of the parameters in sections 2 and 3 the relative sufficiency played, to say the least, an important instrumental role. Following this lead we may ask what conditions are required for the existence of sufficient statistics for the θ, provided the σ are known, and vice versa. From well-known theorems on sufficiency it follows that, apart from a normalizing function depending on θ_ν and σ_i only, $\log P\{x|\theta_\nu, \sigma_i\}$ must be a bilinear function of some transformations θ'_ν and σ'_i of θ_ν and σ_i, respectively. Therefore, on proper choice of θ_ν and σ_i we have

$$P\{x|\theta_\nu, \sigma_i\} = \frac{1}{\gamma(\theta_\nu, \sigma_i)} \exp\,[\varphi(x)\theta_\nu + \psi(x)\sigma_i + \chi(x)\theta_\nu\sigma_i + \rho(x)], \tag{4.2}$$

where φ, ψ, χ, ρ are functions of x only. In writing (4.2), it is presumed that the parameters are one-dimensional, but, in order to make the formula apply to higher dimensional cases, we just have to interpret $\varphi(x)\theta_\nu$ and $\psi(x)\sigma_i$ as inner products of vectors, while $\chi(x)$ is taken to be a matrix and $\chi(x)\theta_\nu\sigma_i$ to be a homogeneous bilinear form in the elements of θ_ν and σ_i. For the sake of simplicity, most of the formulas in the following will be expressed as if θ_ν and σ_i were scalars, but the above reinterpretation is available at any time.

I have not yet studied the general form (4.2) in detail, but the case $\chi(x) = 0$, for which there exist relatively sufficient estimators for the θ which are independent of the σ and vice versa, shows several features of considerable interest.

Preparatory to the analysis of this case I shall change the stochastic variable by introducing the selection vector

$$\mathfrak{a}_{\nu i} = (0, \cdots, 1, \cdots, 0) = (a_{\nu i}^{(1)}, \cdots, a_{\nu i}^{(m)}), \tag{4.3}$$

which has elements 0 except the μth where $x_{\nu i} = x^{(\mu)}$. Writing, furthermore,

$$\varphi_\mu = \varphi(x^{(\mu)}), \qquad \psi_\mu = \psi(x^{(\mu)}), \qquad \rho_\mu = \rho(x^{(\mu)}) \tag{4.4}$$

and introducing the vectors

$$\Phi = (\varphi_1, \cdots, \varphi_m), \qquad \Psi = (\psi_1, \cdots, \psi_m), \qquad \mathbf{P} = (\rho_1, \cdots, \rho_m), \tag{4.5}$$

we have

$$P\{\mathfrak{a}_{\nu i}\} = \frac{1}{\gamma(\theta_\nu, \sigma_i)} \exp\,(\theta_\nu \mathfrak{a}_{\nu i}\Phi^* + \sigma_i \mathfrak{a}_{\nu i}\Psi^* + \mathfrak{a}_{\nu i}\mathbf{P}^*). \tag{4.6}$$

Now imagine that all of the k stimuli, $i = 1, \cdots, k$ have been applied to each of the n persons $\nu = 1, \cdots, n$ and that we obtain a selection vector for a response every time. We may then build up a matrix of order (n, k), whose elements are the selection vectors $\mathfrak{a}_{\nu i}$. For each person the $\mathfrak{a}_{\nu i}$ add to a total vector $\mathfrak{a}_{\nu\cdot}$, and similarly, the $\mathfrak{a}_{\nu i}$ for a given stimulus add to a total vector $\mathfrak{a}_{\cdot i}$. Out of these marginal vectors, whose grand total is $\mathfrak{a}_{\cdot\cdot}$, we form the ordinary matrices

$$A_{*\cdot} = \begin{pmatrix} \mathfrak{a}_{1\cdot} \\ \cdots \\ \mathfrak{a}_{n\cdot} \end{pmatrix}, \qquad A_{\cdot *} = \begin{pmatrix} \mathfrak{a}_{\cdot 1} \\ \cdots \\ \mathfrak{a}_{\cdot k} \end{pmatrix} \tag{4.7}$$

of orders (n, m) and (k, m).

Noticing that

$$\begin{aligned} \sum_\nu \sum_i \theta_\nu \mathfrak{a}_{\nu i} &= \Theta A_{*\cdot}, \\ \sum_\nu \sum_i \sigma_i \mathfrak{a}_{\nu i} &= \Sigma A_{\cdot *} \end{aligned} \tag{4.8}$$

and writing

$$\gamma(\Theta, \Sigma) = \prod_\nu \prod_i \gamma(\theta_\nu, \sigma_i), \tag{4.9}$$

we obtain for the distribution of the whole set of $\mathfrak{a}_{\nu i}$

$$P\{A\} = \frac{\exp\,(\Theta A_{*\cdot}\Phi^* + \Sigma A_{\cdot *}\Psi^* + \mathfrak{a}_{\cdot\cdot}\mathbf{P}^*)}{\gamma(\Theta, \Sigma)}. \tag{4.10}$$

Since $P\{A\}$ remains the same for all matrices $(\mathfrak{a}_{\nu i})$ with the same marginal matrices (4.7), the joint distribution of $A_{*\cdot}$ and $A_{\cdot *}$ becomes

$$P\{A_{*\cdot}, A_{\cdot *}\} = \begin{bmatrix} A_{*\cdot} \\ A_{\cdot *} \end{bmatrix} \frac{\exp\,(\Theta A_{*\cdot}\Phi^* + \Sigma A_{\cdot *}\Psi^* + \mathfrak{a}_{\cdot\cdot}\mathbf{P}^*)}{\gamma(\Theta, \Sigma)} \tag{4.11}$$

where the coefficient on the right denotes the number of such matrices for which the elements are selection vectors. By convention the coefficient is understood to be 0 if the two matrices could not possibly represent the two sets of marginal vectors of the same A-matrix, for example, if some $a_{\nu\cdot} > k$ or some $a_{\cdot i} > n$, or if $\sum_\nu A_{\nu\cdot} \neq \sum_i A_{\cdot i}$.

On summation of (4.11) over all matrices $A_{*\cdot}$ and $A_{\cdot *}$ with the same grand total vector $\mathfrak{a}_{\cdot\cdot}$, we obtain the distribution

$$P\{\mathfrak{a}_{\cdot\cdot}\} = \frac{\gamma_{\mathfrak{a}_{\cdot\cdot}}(\Theta, \Sigma)}{\gamma(\Theta, \Sigma)} e^{\mathfrak{a}_{\cdot\cdot}\mathbf{P}^*}, \tag{4.12}$$

where

$$\gamma_{\mathfrak{a}..}(\Theta, \Sigma) = \sum_{A_{*.}} \sum_{A_{.*}} \begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix} \exp\left(\Theta A_{*.}\Phi^* + \Sigma A_{.*}\Psi^*\right). \tag{4.13}$$

From (4.11) and (4.12) we derive the conditional distribution of $A_{*.}$ and $A_{.*}$ for given $\mathfrak{a}..$,

$$P\{A_{*.}, A_{.*}|\mathfrak{a}..\} = \begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix} \frac{\exp\left(\Theta A_{*.}\Phi^* + \Sigma A_{.*}\Psi^*\right)}{\gamma_{\mathfrak{a}..}(\Theta, \Sigma)} \tag{4.14}$$

in which the terms containing **P** have disappeared. Next we may sum (4.14) over the matrices $A_{.*}$ to find the distribution of $A_{*.}$ for given $\mathfrak{a}..$,

$$P\{A_{*.}|A_{.*}\} = \frac{\gamma_{A_{*.}}(\Sigma)}{\gamma_{\mathfrak{a}..}(\Theta, \Sigma)} e^{\Theta A_{*.}\Phi^*}, \tag{4.15}$$

where

$$\gamma_{A_{*.}}(\Sigma) = \sum_{A_{.*}} \begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix} e^{\Sigma A_{.*}\Psi^*}. \tag{4.16}$$

Now divide (4.15) into (4.14) to obtain the conditional distribution of one marginal matrix, given the other,

$$P\{A_{.*}|A_{*.}\} = \begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix} \frac{e^{\Sigma A_{.*}\Psi^*}}{\gamma_{A_{*.}}(\Sigma)}. \tag{4.17}$$

This distribution is seen to be independent of Θ. This means that *we may estimate the stimulus parameters without regard to the personal parameters, instead of which we use the observed personal marginals.*

Symmetrically, of course, we have

$$P\{A_{.*}|\mathfrak{a}..\} = \frac{\gamma_{A_{.*}}(\Theta)}{\gamma_{\mathfrak{a}..}(\Theta, \Sigma)} e^{\Sigma A_{.*}\Psi^*}, \tag{4.18}$$

where

$$\gamma_{A_{.*}}(\Theta) = \sum_{A_{*.}} \begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix} e^{\Theta A_{*.}\Phi^*}, \tag{4.19}$$

and consequently

$$P\{A_{*.}|A_{.*}\} = \begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix} \frac{e^{\Theta A_{*.}\Phi^*}}{\gamma_{A_{.*}}(\Theta)} \tag{4.20}$$

so that the personal parameters may be estimated without regard to the stimulus parameters, which are replaced by the observed stimulus marginals.

Finally, from (4.10) and (4.11) we obtain the conditional distribution of the entire observed set of selection vectors for given marginal matrices,

$$P\{A|A_{*.}, A_{.*}\} = \frac{1}{\begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix}}. \tag{4.21}$$

Since this distribution is independent of both sets of parameters, *it may serve as a basis for nonparametric checks of our model* (4.2). It should be noticed that all algebraically possible A-matrices are equally probable if the model holds. For an effective check, the $\mathfrak{A}_{\nu i}$ should be collected into groups pointing to possible deviations from the model in a more or less specified direction.

So far, our results run parallel to those obtained for the item analysis model (3.2), but when we try to apply (4.17) and (4.20) for estimation purposes, problems of a new kind arise. Let us try for a moment to think of (4.2) as fully specified, for example, by the Poisson law (2.1) and (2.2), with the functions $\varphi(x)$, $\psi(x)$, and $\rho(x)$ assumed known, and $\chi(x) \equiv 0$. From (4.17) it follows that the vector

$$\mathcal{S}^* = A_{.*}\Psi^* \tag{4.22}$$

is a sufficient statistic for Σ with distribution

$$P\{\mathcal{S}|A_{*.}\} = \sum_{A_{.*}\Psi^*=\mathcal{S}^*} \begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix} \frac{e^{\Sigma \mathcal{S}^*}}{\gamma_{A_{*.}}(\Sigma)} = \begin{Bmatrix} A_{*.} \\ \mathcal{S} \end{Bmatrix} \frac{e^{\Sigma \mathcal{S}^*}}{\gamma_{A_{*.}}(\Sigma)}, \tag{4.23}$$

say, and the corresponding conditional distribution

$$P\{A_{.*}|A_{*.}, \mathcal{S}\} = \frac{\begin{bmatrix} A_{*.} \\ A_{.*} \end{bmatrix}}{\begin{Bmatrix} A_{*.} \\ \mathcal{S} \end{Bmatrix}} \tag{4.24}$$

is independent of Σ.

At this stage two comments are called for. In accordance with ordinary psychometric practice, the possible observations $x^{(1)}, \cdots, x^{(m)}$ might have been translated to a set of scores, $z^{(1)}, \cdots, z^{(m)}$, say. Provided, however, that our model is given by (4.2), or equivalently by (4.6), the values $\psi_1, \cdots, \psi_m$ will be unaffected by the choice of the z. Since

$$s_i = \sum_{\mu} a_{.i}^{(\mu)}\psi_\mu, \qquad i = 1, \cdots, k \tag{4.25}$$

are sufficient for estimating the σ, the ψ_μ represent the only way in which the x should enter the estimation of the σ. *We may, therefore, ignore the z and consider the ψ as the proper scoring function of x so far as the estimation of the stimulus parameters is concerned.*

Symmetrically, of course, $\varphi(x)$ is the proper scoring function for the estimation of the personal parameters. I shall discuss the particular role of $\rho(x)$ on some other occasion.

The other comment refers to (4.24). Due to the independence of Σ, this relation might be employed just as well as (4.21) in a check of the model. There is, however, an essential difference between the kinds of check exerted by the two formulas. Relation (4.21) is independent both of parameters and scoring functions and, in fact, even of the dimensions of the θ and the σ. *Thus, checks based on* (4.21) *refer only to the pure framework;* can the given data be represented at all

by any model whatsoever of the form (4.2) with $\chi(x) \equiv 0$. On the other hand, (4.24) depends on the choice of the function $\psi(x)$, and therefore, if the given data have passed the first check, *the check based on* (4.24) *would be interpreted mainly as a check on the suitability of the choice of the scoring function* $\psi(x)$.

Now it may very well happen that the framework according to the first test seems quite feasible but that the scoring function according to the second test seems quite inadequate. *In general, therefore, we may be better off if we leave the scoring function unspecified altogether, and try to estimate it from the data.*

Accordingly, our problem now is to estimate the $m + k$ elements of Σ and Ψ from an observed matrix $A_{.*}$ of order (m, k). This, of course, is quite possible insofar as (4.17) gives an adequate representation of these data. Estimation and checking may be combined as in the following procedure.

The exponent in (4.17) is a bilinear form in the ψ and the σ,

$$\Sigma A_{.*}\Psi^{*} = \sum_{i}\sum_{\mu} a_{.i}^{(\mu)}\psi_{\mu}\sigma_{i}. \tag{4.26}$$

Consider now for a moment

$$\tau_i^{(\mu)} = \psi_\mu \sigma_i \tag{4.27}$$

as mk algebraically almost independent (apart from a trivial normalization) parameters to be estimated from the observed numbers $a_{.i}^{(\mu)}$. Denote the estimate of $\tau_i^{(\mu)}$ by $t_i^{(\mu)}$,

$$t_i^{(\mu)} \risingdotseq \tau_i^{(\mu)}. \tag{4.28}$$

If σ_i, and therefore also ψ_μ are one-dimensional, $t_i^{(\mu)}$ is an estimate of the product $\psi_\mu\sigma_i$; the joint distribution of these estimates is known in principle. As a consequence, the $t_i^{(\mu)}$ for any fixed i should be "stochastically proportional" to the $t_.^{(\mu)}$-values and, symmetrically, the $t_i^{(\mu)}$ for any fixed μ should be "stochastically proportional" to the $t_i^{(\cdot)}$-values.

In case our data satisfy these conditions, (4.17) would seem applicable with the σ_i and the ψ_μ interpreted as scalars which may be estimated from the $t_i^{(\cdot)}$ and the $t_.^{(\mu)}$. Better estimates may be available but this is a technical matter which I shall leave aside on this occasion.

However, it may equally well happen that the proportionality condition clearly fails to hold. Then the conclusion is that the assumption now at stake, namely, the one-dimensionality of the σ and the ψ, has to be dropped. If so, the alternative assumption obviously is that the parameters and the scoring function may be of a higher dimension. For an investigation of this possibility we turn to the reinterpretation of our formulas. Accordingly putting

$$\sigma_i = (\sigma_{i1}, \cdots, \sigma_{il}), \qquad \psi_\mu = (\psi_{\mu 1}, \cdots, \psi_{\mu l}), \tag{4.29}$$

we have to interpret $\tau_i^{(\mu)}$ in (4.27) as the inner product of σ_i and ψ_μ. Since $t_i^{(\mu)}$ still estimates $\tau_i^{(\mu)}$, we now have

$$t_i^{(\mu)} = \sum_{\lambda=1}^{l} \sigma_{i\lambda}\psi_{\mu\lambda} + u_i^{(\mu)}, \tag{4.30}$$

where the residuals $u_i^{(\mu)}$ under favorable conditions may be fairly small and even approximately normally correlated.

Formally, the system of relations (4.30) *reminds us strongly of factor analysis specification, from which, however, it differs in a main respect in that the joint distribution of the* $t_i^{(\mu)}$ *is in principle known and depends only on the basic parameters* (4.29). In particular, the variations of the residuals are not to be accounted for by a particular set of variability parameters, and assumptions about variances and covariances of the residuals that are not corollaries from the distribution of the $t_i^{(\mu)}$ would be inadmissible.

Another consequence of the distribution of the $t_i^{(\mu)}$ is that the obvious analogue of an old and troublesome question in factor analysis, "When to stop factoring?", here has a fairly definite answer insofar as it would seem pointless to try to add another pair of elements to (4.29) when the calculated residuals after l pairs already tally with the distribution of the $u_i^{(\mu)}$, as computed from the l-dimensional estimates of the σ and the ψ.

On following the same line of deductions from (4.20), we eventually arrive at a complete mastery of the estimation of both sets of parameters and of both scoring functions, including the dimensionalities. With this, finally we may estimate the third scoring function $\rho(x)$ from $\mathfrak{A}_{..}$ on the basis of (4.12). In a certain sense, $\rho(x)$ gives the final specification of the model (4.6), the exponential framework having been agreed upon.

5. Principles of comparison

In psychology proper, and in particular in its applications, there seems to be a strong need for replacing the original qualitative observations by measurable quantities. In formulating a general law of the type (4.2), we have so far replaced the observations by quantitative parameters, but that does not imply that we have a proper measurement, on a ratio scale or on an interval scale, of the individuals or of the stimuli nor even that a proper ordering is available. This is obvious in case the parameters are nonscalar since the relations $<$ and $>$ seem hard to extend to higher dimensions. But even in case of scalar parameters, as we shall see, trouble may arise when m exceeds 2.

In an attempt to make clear what may be achieved in these respects, within the class of models considered here, I shall begin with simply dropping both measuring and ordering as possibly too ambitious concepts. More modestly I shall inquire into the possibilities for *just comparing individuals and comparing stimuli.* In doing so, I shall, however, formulate four requirements that to my mind seem indispensable for well-defined comparisons. Preliminarily, it may be noted that in order to compare stimuli we have to apply them to some adequately chosen individuals, and similarly, that in order to compare individuals in a given respect, we must use some adequate stimuli.

Now the requirements are as follows.

The comparison between two stimuli should be independent of which particular

individuals were instrumental for the comparison; and it should also be independent of which other stimuli within the considered class were or might also have been compared.

Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered were instrumental for the comparison; and it should also be independent of which other individuals were also compared, on the same or on some other occasion.

Returning now to (4.1) we shall obviously be in a favorable position if this equation, whatever ξ may be, has a unique solution with regard to θ for any given σ, and vice versa, that is, if the functions

$$\theta = \lambda(\sigma, \xi) \quad \text{and} \quad \sigma = \kappa(\theta, \xi) \tag{5.1}$$

are uniquely defined.

In this situation we may compare two individuals by means of the following *principle of equivalent stimuli:* To any chosen ξ, which, it will be recalled, is not a response but a distribution parameter, we may for two individuals with the parameters θ and θ' find the stimuli σ and σ' that produce the reaction ξ. Here σ and σ' are called equivalent stimuli for the comparison of θ and θ'. *Now letting ξ vary throughout its entire range, we obtain a series of equivalent pairs of stimuli, the relationship between which constitutes a well-defined comparison between θ and θ'.* In fact, for any stimulus σ as applied to the individual θ we may look up which stimulus σ' corresponds to it in the individual θ'.

This comparison fulfills the fourth of our requirements. In fact, considering a third individual with the parameter θ'' it is easy to see that *if the first comparison is followed by a comparison of θ'' to θ' we shall get the same σ'' as would have obtained by direct comparison of θ'' to θ.*

Similarly, we may compare two stimuli by a *principle of equivalent individuals:* To any chosen ξ we may for the two stimuli considered, with parameters σ and σ' find the two individuals, with parameters θ and θ', which react with ξ upon σ and σ'. *The relationship between all such equivalent individuals constitutes a well-defined comparison between the stimuli considered.* Obviously *such comparisons are also transitive,* thus fulfilling the second of our requirements.

On the other hand, if equation (4.1) for some θ or for some σ has more than one solution, then the transitivity cannot hold without exception. And if the equation for some θ or for some σ has no solution corresponding to a given ξ, then a comparison cannot always be carried out.

Thus, the existence and the uniqueness of the functions (5.1) *are necessary and sufficient conditions for unrestricted and transitive comparability of both individuals and stimuli.*

The rule of transitivity seems to generalize one of the most fundamental properties of measurement. If, for instance, we wish to measure the distance between two points A and C on a straight line we may do it directly or we may interpose a third point B, measure the distance AB, and on top of that measure the distance BC to obtain the total AC.

6. On measurement

As a corollary we may conclude that *in the case of transitivity the dimensions of ξ, θ, and σ must be equal.*

For the model (4.2) this implies that $\psi(x)$ must be a linear transform of $\varphi(x)$, which is always so in the trivial case $l \geqq m$ but has to be tested in the case $l < m$. Equivalently, this implies that with an appropriate normalization of the σ we have

$$\psi(x) \equiv \varphi(x), \tag{6.1}$$

in which case the model simplifies to

$$P\{x|\xi_{\nu i}\} = \frac{\exp\left[\xi_{\nu i}\varphi^*(x) + \rho(x)\right]}{\gamma(\theta_\nu, \sigma_i)}, \tag{6.2}$$

where

$$\xi_{\nu i} = \theta_\nu + \sigma_i. \tag{6.3}$$

According to this formula, which corresponds to the logarithmic version of (2.2), we may introduce what may be called *the l-dimensional measurement of the personal and the stimulus parameters,* choosing, for example, a certain stimulus, $i = 0$ say, as our reference point with

$$\sigma_0 = (0, \cdots, 0). \tag{6.4}$$

With such a choice all other parameters θ_ν and σ_i are uniquely determined by virtue of the empirically established law (6.3). The parallel between this type of approach and the introduction of mass and force as measurable concepts in classical dynamics has been thoroughly discussed in chapter 7 of [3].

Accordingly, the first and the third requirements also are fulfilled in the model (6.2).

A discussion of these requirements as separated from (4.2) awaits another occasion.

REFERENCES

[1] B. F. Skinner, "A case history in scientific method," *Amer. Psych.*, Vol. 11 (1956), pp. 221–233.

[2] J. Zubin (ed.), *Experimental Abnormal Psychology*, New York, Columbia University Store, 1955, pp. 2–28. (Mimeo.)

[3] G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen, The Danish Institute of Educational Research, 1960.